Michael, A. J., & Werner, M. J. (2018). Preface to the focus section on the collaboratory for the study of earthquake predictability (CSEP): New results and future directions. *Seismological Research Letters*, *89*(4), 1226-1228. https://doi.org/10.1785/0220180161

Preface to the Focus Section on the Collaboratory for the Study of Earthquake Predictability (CSEP): New Results and Future Directions

Andrew J. Michael
USGS
345 Middlefield Road
Menlo Park, CA 94025 USA
ajmichael@usgs.gov

Maximilian J. Werner
School of Earth Sciences & Cabot Institute
University of Bristol, Bristol, UK
max.werner@bristol.ac.uk

The Collaboratory for the Study of Earthquake Predictability (CSEP; Jordan, 2006) carries out fully prospective tests of earthquake forecasts, using fixed and standardized statistical tests and authoritative data sets, to assess the predictive skill of forecast models and to make objective comparisons between models. CSEP conducts prospective experiments at four testing centers around the world, at which over four hundred models and model versions are currently under evaluation. These models include a range of methods and scales from long-term global earthquake forecasts to short-term regional forecasts used for Operational Earthquake Forecasting (OEF). CSEP has also conducted retrospective tests and developed new testing methods in its quest to answer fundamental scientific questions, improve seismic hazard assessments, and develop new forecast methods for OEF.

This Focus Section presents a sample of the work being done within CSEP testing centers (cseptesting.org) and studies that use CSEP methods outside of the testing centers to fulfill these goals at a crucial moment in the development of this collaboration as it sets new goals for a second phase. Building on the insights of the first phase, CSEP is planning to expand its scope by developing procedures for testing different types of forecasts of the earthquake process, designing new tests with fewer assumptions that can also address epistemic uncertainties, and providing a more flexible software system. Thus, the goals of this Focus Section are to inform the broader seismological community about CSEP activities and findings, and to stimulate new ideas and involvement as CSEP moves forward.

As an example of work using the prospective CSEP framework, Rhoades et al. (2018) discuss results from the New Zealand testing center at GNS Science from 2008 to 2017. That period includes multiple major earthquakes such as the 2016 M7.8 Kaikōura earthquake and the Canterbury sequence, which started with the 2010 M7.1 Darfield earthquake and includes the devastating 2011 M6.2 Christchurch earthquake. Such an active time period allows them to demonstrate the value of standard clustering models that include Omori-Utsu time decay, the EEPAS (Every Earthquake a Precursor According to Scale) model, which does not include that time-dependent decay term, and the value of updating spatially smoothed seismicity models. They conclude that having the CSEP tests underway assisted in the development of hybrid time-varying models for public real-time OEF during these major earthquake sequences.

In encouraging news for physics-based forecasting models, Cattania et al. (2018) describe new CSEP results from a retrospective experiment during the 2010-2012 Canterbury, New Zealand, earthquake sequence that show substantial improvements by recently developed Coulomb-based models. The static Coulomb stress hypothesis is perhaps the most widely accepted physical mechanism for (near-field) earthquake triggering, but its predictive power remains controversial (e.g. Woessner et al., 2011). The Canterbury experiment was designed as part of the European FP7 project REAKT (http://www.reaktproject.eu/) to assess the ability of a variety of new Coulomb-based, statistical and hybrid models to forecast a complex earthquake cascade with 1-day, 1-month and 1-year forecast horizons. Fourteen models were developed by researchers from around the world and their source codes installed at the SCEC testing center. The retrospective nature of the experiment enabled modelers to make use of data products that are not (yet) routinely available in real-time, such as authoritative finite fault and slip models. Cattania et al. conclude that new Coulomb/rate-state models that account for uncertainties and secondary triggering can compete with standard statistical clustering models (e.g. ETAS, STEP). However, the most skillful forecasts were generated by statistical models that exploit the geometry of the ruptured fault. In addition, ETAS and STEP models fared marginally better than new hybrid versions that include Coulomb-based spatial aftershock footprints. Nonetheless, Cattania's results are encouraging for physics-based forecasting, and may point the way for further model development.

Transitioning back to prospective CSEP results, Taroni et al. (2018) discuss 1-day, 3-month, and 5-year forecasts for Italy where CSEP tests have been underway since 2009 as part of the European testing center at ETH Zurich. They found that 3-month forecasts provided little additional value to 5-year time-independent forecasts but did find utility in 1-day ETAS forecasts, whose smoother spatial aftershock decay than that of STEP forecasts provided better forecasts during sequences. Only 14 earthquakes M>4.95 occurred since 2009, but these indicate that 5-year time-independent models that combine smoothed historical and instrumental seismicity with fault-based information outperform other models. For the 1-day forecasts, they demonstrate the value of ensemble models that combine multiple forecasts. These ensemble models perform about as well as the best models, but have greater stability than a single best model chosen at the start of a forecasting period. These results have been used as the basis for Italy's 1-week OEF system (Marzocchi et al., 2014).

Strader et al. (2018) discuss global CSEP tests of the GEAR1 model (Bird et al., 2015) and the smoothed seismicity and deformation models that are its ingredients. The prospective test, as opposed to model development, started in 2015 and the test uses events with M≥5.95. This study used both the original CSEP likelihood tests, which assess whether the observations are consistent with the model (Zechar et al., 2010), and model comparison tests based on relative informativeness (Rhoades et al, 2011). Despite the brief time period, their results suggest that the GEAR1 ensemble model is superior to its individual ingredients because the deformation model provides a broader zone of earthquake activity than observed in the short period of seismicity observations.

Bird (2018) independently tested these models by calculating the information gains for three independent, prospective years: 2014, 2015, and 2016, and using a lower magnitude threshold (5.767) for consistency with how the models were developed. He reached similar conclusions to Strader et al. showing the stability of these tests over both short periods of time and variations in the magnitude threshold. These global tests support including deformation (e.g., horizontal strain-rates) in seismic hazards assessments, as was done in Uniform California Earthquake Rupture Forecast v3 (UCERF3, Field et al., 2014).

Akinci et al. (2018) use CSEP methods to develop and assess a suite of smoothed seismicity models as part of the effort to make the 2017-2018 update to the Italian Probabilistic Seismic Hazard Maps (IPSHA). Previously, the IPSHA used source zones determined from expert opinion. Instead, Akinci et al. use likelihood tests to optimize both fixed kernel and adaptive smoothing methods and combine these methods with both historical and instrumental catalogs to produce an ensemble earthquake rate model for Italy, estimate its uncertainty, and compare this new model to those already under test in CSEP's Italy region. The authors argue that using the CSEP tests and ensemble modeling approaches developed as part of CSEP results in a more objective and better performing model for the IPSHA. This highlights a key advantage of CSEP: its common regions and predefined datasets provide for reproducible comparisons and benchmarking for new models.

The work by Akinci et al. mirrors other uses of CSEP tests to improve PSHA products. In the U.S., the National Seismic Hazard Maps (NSHM) have used CSEP testing methods to optimize spatial smoothing for both the 50-year national hazard maps (Petersen et al., 2014) and the 1-year hazard maps that focus on induced seismicity in the Central and Eastern U.S. (Petersen et al., 2016). The results of the RELM experiment (Field, 2007), which became the first experiment conducted within CSEP, were used to select an adaptive smoothing method (Helmstetter et al., 2007) for use as one logic tree branch in UCERF3 (Field et al., 2014). UCERF3 is used as the earthquake rate model for California in the 50-year NSHM, thus highlighting how CSEP results increasingly affect hazard products and make an impact outside the research community.

Testing long-term seismic hazards assessments often seems impossible due to the long time periods involved. Jackson (2018) points out that the 30-year forecast released by the Working Group on California Earthquake Probabilities in 1988 (WGCEP, 1988) is now virtually over. This fault based forecast does not fit neatly into the current CSEP forecast framework and Jackson admits that his assessment is not truly a prospective test because the testing methods were not agreed upon when the forecast was published. Still, he applies CSEP number tests and spatial tests to the 16 fault segments. During the forecast period only one earthquake is clearly associated with the forecast (the M6 Parkfield earthquake in 2004) and one other could be associated with the forecast (the M6.8 Loma Prieta in 1989), though it isn't clear whether that event is truly on the San Andreas fault (Harris, 1998). Only if the Loma Prieta earthquake is included does the forecast pass the CSEP number test. Jackson suggests a simpler, time-independent alternative model as a reference model that turns out to outperform the time-dependent WGCEP88 model. Although this is not a rigorous, prospective test, Jackson shows that thinking through the tests illustrates how to make future forecasts testable and emphasizes the difficulty of testing forecasts of high magnitude earthquakes over limited regions.

For both long-term forecasts and short-term forecasts during aftershock sequences, Ogata et al. (2018) examine whether spatial and/or temporal variations in the b-value improve the forecasts over using a constant value of b=0.9 as is now the standard practice in Japan. Using the Information gain criteria to study a variety of methods over multiple earthquake sequences, they do not find a clear advantage to using a variable b-value. They do, however, suggest that other methods to estimate variations may still prove valuable and argue that it is critical to accurately estimate the magnitude-frequency distribution due to the importance of forecasting the largest and most-damaging but rarest aftershocks.

Schorlemmer et al. (2018) summarize CSEP's philosophy, activities and achievements beyond the limited scope of the other eight papers in this Focus Section. They distill insights into earthquake predictability from a decade of CSEP operations, and present ideas for future CSEP activities. The future directions under development include tests of fault-based forecasts that include finite fault information rather than just testing the locations of hypocenters, simulation-based forecasts and tests that better approximate earthquake clustering, and tests of ground motion measures to directly assess probabilistic seismic hazards assessments.

We hope that the articles in this Focus Section provide a window into CSEP's activities and capabilities, inspire the readers to learn more, use CSEP's testing methods, and participate in our global CSEP community. The next phase of CSEP will require continued international collaboration, rely on global, high-quality data sets, and draw on new forecasting approaches to help CSEP achieve its goals of improving earthquake hazard assessments and providing insights into fundamental questions about earthquake behavior.

Acknowledgements

We thank Editor-in-Chief Zhigang Peng and Managing Editor Mary George for their guidance during the preparation of this Focus Section, and both the authors and reviewers for thoughtful and speedy work. We appreciate thoughtful reviews by Morgan Page and Michael Blanpied.

References

Akinci, A., M.P. Moschetti, and M. Taroni (2018). Ensemble smoothed seismicity models for the new Italian probabilistic seismic hazard map, *Seismol. Res. Lett.* (this issue).

Bird, P. (2018). Ranking some global forecasts with the Kagan information score, *Seismol. Res. Lett.* (this issue).

Cattania C., M. J. Werner, W. Marzocch, S. Hainzl, D. Rhoades, M. Gerstenberger, M. Liukis, W. Savran, A. Christophersen, A. Helmstetter, A. Jimenez, S. Steacy, and T. H. Jordan (2018). The forecasting skill of physics-based seismicity models during the 2010-2012 Canterbury, New Zealand, earthquake sequence, *Seismol. Res. Lett.* (this issue).

Field, E. H. (2007). Overview of the working group for the development of Regional Earthquake Likelihood Models (RELM), Seismol. Res. Lett. 78, no. 1, 7–16, doi: 10.1785/gssrl.78.1.7.

Field, E. H., R. J. Arrowsmith, G. P. Biasi, P. Bird, T. E. Dawson, K. R. Felzer, D. D. Jackson, K. M. Johnson, T. H. Jordan, C. Madden, A. J. Michael, K. R. Milner, M. T. Page, T. Parsons, P. M. Powers, B. E. Shaw, W. R. Thatcher, R. J. Weldon, and Y. Zeng (2014). Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3)--The Time-Independent Model, Bull. Seism. Soc. Am., 104, 1122-1180 doi:10.1785/0120130164.

Harris, R. A. (1998). Forecasts of the 1989 Loma Prieta, California, Earthquake, *Bull. Seism. Soc. Am.* **88,** 898–916.

Helmstetter, A., Y. Y. Kagan and D.D. Jackson (2007). High-resolution time-independent grid based forecast for M≥5 earthquakes in California, *Seismol. Res. Lett.* **78** 78-86, 10.1785/gssrl.78.1.78.

Jackson, D.D. (2018). Testing the classic 1988 forecast, *Seismol. Res. Lett.* (this issue).

Jordan, T. H. (2006). Earthquake predictability, brick by brick, Seism. Res. Lett., 77, 3-6, doi: 10.1785/gssrl.77.1.3.

Marzocchi, W., A. M. Lombardi, and E. Casarotti (2014). The Establishment of an Operational Earthquake Forecasting System in Italy, *Seism. Res. Lett.*, 85, 961-969 doi:10.1785/0220130219.

Ogata Y., K. Katsura, H. Tsuruoka, and N. Hirata (2018). Exploring magnitude forecasting of the next earthquake, *Seismol. Res. Lett.* (this issue).

Petersen, M. D., M. P. Moschetti, P. M. Powers, C. S. Mueller, K. M. Haller, A. D. Frankel, Y. Zeng, S. Rezaeian, S. C. Harmsen, O. S. Boyd, N. Field, R. Chen, K. S. Rukstales, N. Luco, R. L. Wheeler, R. A. Williams, and A. H. Olsen (2014). Documentation for the 2014 update of the United States national seismic hazard maps doi:10.3133/ofr20141091.

Petersen, M. D., C. S. Mueller, M. P. Moschetti, S. M. Hoover, A. L. Llenos, W. L. Ellsworth, A. J. Michael, J. L. Rubinstein, A. F. McGarr, and K. S. Rukstales (2016). Seismic-Hazard Forecast for 2016 Including Induced and Natural Earthquakes in the Central and Eastern United States, Seism. Res. Lett., 87, 1327-1341 doi:10.1785/0220160072.

Rhoades, D. A., A. Christophersen, M.C. Gerstenberger, M. Liukis, F. Silva, W. Marzocchi, M.J. Werner, and T.H. Jordan (2018). Highlights from the first ten years of the New Zealand earthquake forecast testing center, *Seismol. Res. Lett.* (this issue).

Rhoades, D.A., Schorlemmer, D., Gerstenberger, M.C. et al. Acta Geophys. (2011) 59: 728. https://doi.org/10.2478/s11600-011-0013-5.

Schorlemmer, D., M. Gerstenberger, S. Wiemer, D. D. Jackson, and D. A. Rhoades (2007). Earthquake likelihood model testing, Seism. Res. Lett., 78, 17-29, doi: 10.1785/gssrl.78.1.17.

Schorlemmer D., M. J. Werner, W. Marzocchi, T. H. Jordan, Y. Ogata, D. D. Jackson, S. Mak, D. A. Rhoades, M. C. Gerstenberger, N. Hirata, M. Liukis, P. J. Maechling, A. Strader, M. Taroni, S. Wiemer, J. D. Zechar, and J. Zhuang (2018). The Collaboratory for the Study of Earthquake Predictability: achievements and priorities, *Seismol. Res. Lett.* (this issue).

Strader A., M. Werner, J. Bayona, P. Maechling, F. Silva, M. Liukis, and D. Schorlemmer (2018). Prospective evaluation of global earthquake forecast models: two years of observations provide preliminary support for merging smoothed seismicity with geodetic strain rates, *Seismol. Res. Lett.* (this issue).

M. Taroni, W. Marzocchi, D. Schorlemmer, M. J. Werner, S. Wiemer, J.D. Zechar, L. Heiniger, F. Euchner (2018). Prospective CSEP evaluation of 1-day, 3-month, and 5-year earthquake forecasts for Italy, *Seismol. Res. Lett.* (this issue).

Woessner, J., Hainzl, S., Marzocchi, W., Werner, M. J., Lombardi, A. M., Catalli, F., ... & Wiemer, S. (2011). A retrospective comparative forecast test on the 1992 Landers sequence. *Journal of Geophysical Research: Solid Earth*, *116*(B5).

Working Group on California Earthquake Probabilities (WGCEP) (1988). Probabilities of large earthquakes occurring in California on the San Andreas fault, *U.S. Geol. Surv. Open-File Rept. 88-398*, 62 pp.

Zechar, J. D., M. C. Gerstenberger, and D. A. Rhoades (2010). Likelihood-Based Tests for Evaluating Space-Rate-Magnitude Earthquake Forecasts, Bull. Seism. Soc. Am., 100, 1184-1195 doi:10.1785/0120090192.